Leading Edge
Minireview
Cell

Linking Cohesin to Gene Regulation

Daniel Peric-Hupkes[1] and Bas van Steensel[1,*]
[1]Division of Molecular Biology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands
*Correspondence: b.v.steensel@nki.nl
DOI 10.1016/j.cell.2008.03.001

During cell division the cohesin complex mediates the pairing of sister chromatids. Emerging evidence shows that cohesin also has roles in interphase cells. New studies, including that of Gullerova and Proudfoot (2008) in this issue, reveal how cohesin is targeted to specific sites on chromosomes and implicate cohesin in the regulation of gene expression.

Cohesin is a protein complex that keeps sister chromatids together from the time of replication in S phase until their separation at the onset of anaphase, a process that is essential for the correct segregation of chromosomes during mitosis. As expected from this essential function in mitosis, cohesin is highly conserved among eukaryotes. The cohesin core complex consists of two very long protein molecules known as SMC1 and SMC3 (structural *m*aintenance of *c*hromosomes) and two smaller subunits termed Scc1/Rad21and Scc3/SA. Together, these proteins form an unusually large ring-shaped structure of about 30–40 nm in diameter. Structural and functional studies strongly suggest that the cohesin ring can encircle DNA. Several models have been proposed for the mechanism by which cohesin rings may hold the two sister chromatids together, but the precise topology is still unclear (Losada, 2007). Cleavage of the Scc1/Rad21 subunit at the onset of anaphase releases cohesin from chromosomes and allows for chromatid segregation. The cohesin core complex interacts with several other proteins (Dorsett, 2007), including a Scc2/Scc4 loading complex that is required for deposition of cohesin onto the sister chromatids.

Several earlier observations suggested that cohesin may have an additional role as a regulator of gene expression during interphase (reviewed in Dorsett, 2007). In the budding yeast *Saccharomyces cerevisiae*, the cohesin subunits SMC1 and SMC3 are required to prevent spreading of heterochromatin from the silenced HMR locus. Mutations in *Drosophila* Nipped-B, the ortholog of the loading complex component Scc2, can affect enhancer-promoter interactions for the *cut* and *Ubx* genes, and two cohesin core complex subunits control *Runx* gene expression in zebrafish (Horsfield et al., 2007). In humans, a developmental disorder called Cornelia de Lange syndrome (CdLS) is frequently caused by mutations in Scc2 (also known as NIPBL in humans) or cohesin subunits (Dorsett, 2007). Cells from CdLS patients do not show detectable defects in sister chromatid cohesion, suggesting an additional role for cohesin.

A series of recent studies have yielded exciting new insights into the targeting mechanisms and gene regulatory functions of cohesin in a variety of species. Mechanistic studies in fission yeast suggest a role in transcription termination in the G2 phase of the cell cycle. In the fruit fly, evidence directly implicates cohesin in regulation of gene expression independent of sister chromatid cohesion. Finally, in mammals cohesin was found to control gene expression through a functional interaction with the insulator protein CTCF. These results show that cohesin is a remarkably versatile complex with diverse functional roles and various targeting mechanisms in different eukaryotes.

Cohesin Targeting and Gene Regulation in Yeast

Detailed mapping by chromatin immunoprecipitation (ChIP) in *S. cerevisiae* and the fission yeast *S. pombe* has revealed that cohesin is positioned along chromosomes in a focal pattern, with an average spacing of roughly 10 kb (Glynn et al., 2004; Lengronne et al., 2004). This distribution is highly nonrandom. In G2 phase, nearly 90% of detectable cohesin is located between two convergently transcribed genes. In contrast, in late G1 phase cohesin is found at different locations. Subsequently, there is a progressive relocation from the presumed loading sites to regions between convergent genes. Interestingly, this relocation is dependent on transcription: at several inactive genes, cohesin is spread along the transcription unit, but it shifts to the downstream regions of these genes upon activation of transcription. Thus, cohesin is cleared from active transcription units and accumulates in regions between convergent genes (Figure 1A). These observations have raised the intriguing possibility that the ring-shaped cohesin complexes are pushed along the DNA fiber by the elongating RNA polymerase and as a consequence accumulate between convergent genes. However, the ring may also be disassembled (Bausch et al., 2007), and it is somewhat puzzling that no enrichment of cohesin can be observed downstream of active genes that are not in a convergent configuration. Another question is whether cohesin is just passively "set aside" between convergent genes because these regions are of little importance, or whether this peculiar positioning of cohesin may have additional functional relevance.

In this issue, Gullerova and Proudfoot (2008) provide new insights into both the molecular mechanism and a function of cohesin accumulation between convergent genes. The study focuses on the fission yeast *S. pombe*, which has the same conspicuous genomic distribution of cohesion-binding sites as *S. cerevisae* (Lengronne et al., 2004). In several pairs of convergently oriented genes, a high degree of readthrough transcription was observed, with transcription of one gene progressing far into the other gene and vice versa. Remarkably, this readthrough transcription is only detectable in the G1 phase of the cell cycle, whereas during G2 transcription

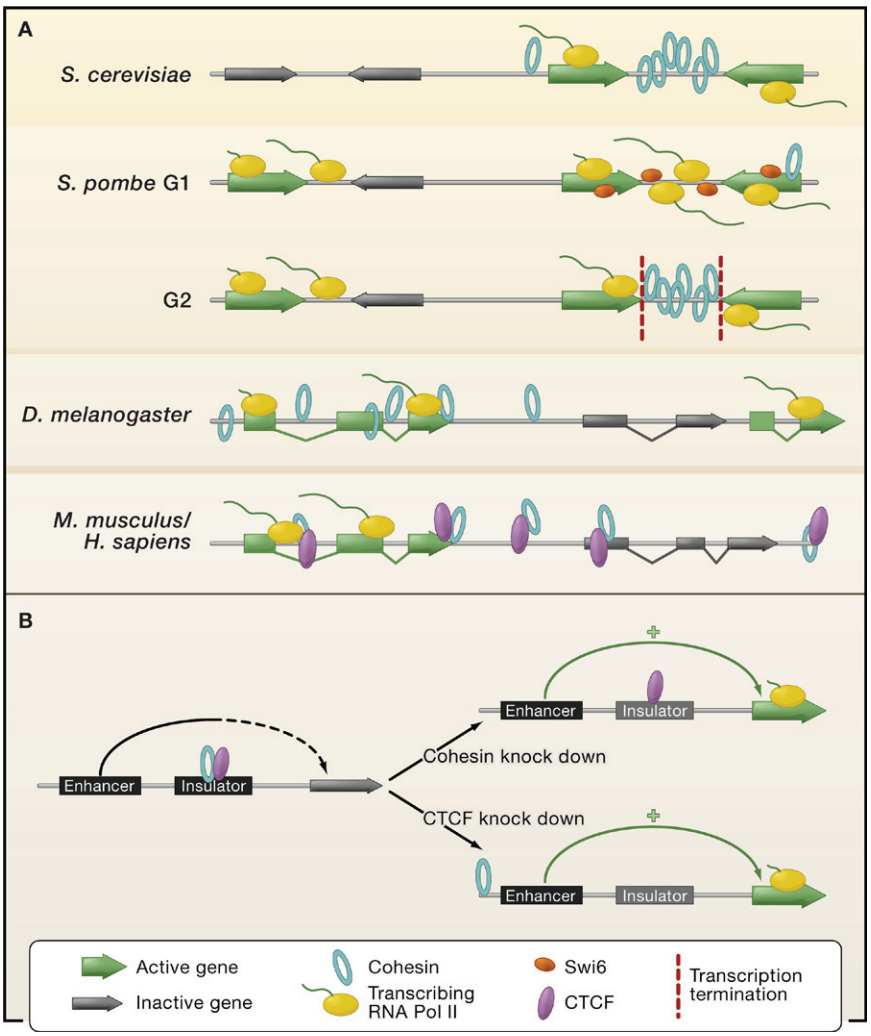

Figure 1. Cohesin Distribution and Gene Regulation

(A) The genomic distributions of cohesin in various eukaryotes are depicted. In *S. cerevisiae* cohesin associates with regions located between convergently transcribed genes. During G1 in *S. pombe* readthrough transcription at convergent genes leads to H3K9 methylation and recruitment of Swi6. In G2 these heterochromatin marks are mostly lost; cohesin accumulates at sites of convergent transcription in G2 and mediates transcription termination. In *D. melanogaster* cohesin is found both in genic and intergenic regions. The population that is bound to genic regions is preferentially located at a subset of active genes and depleted from inactive genes. In mammalian cells cohesin shows a high degree of colocalization with the insulator protein CTCF.

(B) Removal of either cohesin or CTCF causes a loss of CTCF-mediated insulator activity, resulting in increased enhancer-promoter interaction. Cohesin is shown as turquoise rings; it is not known whether in each case the cohesin ring embraces chromatin or associates through a different mechanism.

of the genes is properly terminated. Paradoxically, this correct termination during G2 requires that both genes are active. As reported for budding yeast (Glynn et al., 2004; Lengronne et al., 2004), cohesin accumulates between convergent active genes primarily during G2. This temporal correlation raised the possibility that cohesin helps to terminate transcription. Indeed, deletion of Scc1/Rad21, an essential cohesin subunit, caused a dramatic increase in readthrough transcription of the convergent genes during G2. Thus, cohesin plays a role in the control of transcription termination of convergent genes during G2.

The study of Gullerova and Proudfoot also reveals a mechanism by which cohesin complexes may be targeted to regions between convergently transcribed genes. The readthrough transcription of these genes during G1 may be expected to cause the formation of double-stranded RNAs (dsRNA), which activate the RNA interference (RNAi) pathway, which in turn can target local heterochromatin formation. Indeed, at two pairs of convergently transcribed genes the classical marks of heterochromatin, trimethylation of lysine 9 on histone H3 (H3K9me3) and binding of the Swi6 protein, could be detected between the genes. Deletion of Swi6 caused subsequent loss of cohesin binding, consistent with earlier observations that Swi6 is needed to recruit cohesin to centromeric regions (Bernard et al., 2001; Nonaka et al., 2002). Furthermore, removal of key components of the RNAi pathway led to readthrough transcription during G2. These results suggest a cycle of events involving initial formation of heterochromatin during G1 between convergent genes, which subsequently causes accumulation of cohesin and thereby prevents readthrough transcription during G2 (Figure 1A). What might be the benefit of this intricate cycle, with readthrough transcription being tolerated during G1 but not during G2? It is possible that this provides a means for the cell-cycle regulation of proteins encoded by these genes because differences in 3′ untranslated regions of transcripts can affect mRNA stability and translation efficiency.

A puzzling aspect of the model suggested by Gullerova and Proudfoot is that heterochromatin components are primarily detected at the convergently transcribed regions in G1 phase, whereas cohesin only accumulates at these sites in G2 (Figure 1A). Thus, heterochromatin seems to "pave the way" for cohesin rather than directly recruit cohesin at these loci. Global mapping of cohesin and heterochromatin components in G1 and G2 cells should reveal whether the role of heterochromatin in cohesin targeting represents a general mechanism. In contrast to fission yeast, budding yeast lacks the RNAi and H3K9me3/Swi6 systems. It is possible that an alternative mechanism exists in *S. cerevisiae* to target cohesin to regions between convergently transcribed genes. Sir proteins, which are a different type of heterochromatin proteins, mediate cohesin targeting to the HMR locus in budding yeast (Chang et al., 2005) and might therefore be good candidates to mediate the accumulation of cohesin at sites of convergent transcription.

Cohesin and Gene Regulation in *Drosophila*

Whereas the role of cohesin in gene regulation in *S. pombe* appears tightly linked to the cell cycle, new evidence in *Drosophila* indicates that cohesin can also control transcription in noncycling cells (Pauli et al., 2008; Schuldiner et al., 2008). Cohesin subunits are expressed in postmitotic neurons and disruption of cohesin leads to defects in neuronal patterning that may be caused by altered gene expression. A transposon insertion screen to find mutations that affect axon patterning of mushroom body γ neurons identified two cohesin subunits, SMC1 and Scc3/SA (Schuldiner et al., 2008). Re-expression of SMC1 protein in postmitotic neurons lacking SMC1 could rescue this axon patterning phenotype. Loss of SMC1 causes decreased expression of the ecdysone receptor gene *EcR-B1*, and overexpression of EcR-B1 protein can partially rescue the neuronal defects, indicating that SMC1 may control neuronal patterning via the regulation of EcR-B1 expression. A role for SMC1 could also be demonstrated in dendrite targeting of olfactory neurons. These data directly implicate cohesin in gene regulation in postmitotic cells. Pauli et al. (2008) came to a similar conclusion by replacing the endogenous Scc1/Rad21 subunit with a modified version that is cleavable by tobacco etch virus (TEV) protease. By controlled expression of the protease, the cohesin ring could be disrupted at specific developmental times or in specific tissues. Using this system they show that cohesin is lost from polytene chromosomes within 4 hr upon expression of the TEV protease, but that neither the polytene morphology nor the localization of a wide variety of chromatin proteins is affected. By driving the expression of TEV protease specifically in postmitotic γ neurons or cholinergic neurons, the disruption of cohesin is shown to lead to severe developmental defects of these neuronal subpopulations.

Given these strong indications that *Drosophila* cohesin can regulate gene expression, where is it located on chromosomes? Genome-wide ChIP maps in three different *Drosophila* cell lines (Misulovin et al., 2008) indicate that the pattern of cohesin on *Drosophila* chromosomes is very different from that in yeast. No accumulation is observed between convergent genes. Instead, cohesin shows a broad distribution with preference for transcription units, particularly at intronic sequences and 5′ untranslated regions. Moreover, cohesin overlaps significantly with a subset of transcriptionally active genes and is depleted from inactive genes (Figure 1A). For example, cohesin associates with the *Adb-B* gene in Sg4 cells where the gene is expressed but not in Kc and Bg3 cells in which the gene is inactive. This is in striking contrast to yeast, where cohesin is removed from genes upon transcription. Another difference from yeast is that in *Drosophila* the loading factor Nipped-B displays a very similar distribution to cohesin itself, suggesting that the two components do not dissociate after loading of cohesin onto the DNA. These data indicate that very different targeting mechanisms may govern the locations of cohesin in yeast and flies, which may be related to different roles in gene regulation.

Cohesin and CTCF Function in Mammalian Cells

Three recent papers identify a targeting mechanism and a surprising regulatory role for cohesin in mammals (Parelho et al., 2008; Stedman et al., 2008; Wendt et al., 2008). ChIP maps of several cohesin subunits identified thousands of cohesion-binding sites scattered over the human and mouse genomes, with a slight overrepresentation in coding regions and nearby up- and downstream regions. This pattern was very similar in G1 and G2 synchronized human cells (Wendt et al., 2008). Cohesin-binding sites are detected in many active genes, indicating that (as in *Drosophila*) the transcription machinery does not remove cohesin, at least not permanently (Parelho et al., 2008; Wendt et al., 2008).

Surprisingly, a strong similarity was found between the genomic distribution of cohesin and that of CTCF, a sequence-specific DNA-binding protein that has been implicated as a transcriptional regulator, insulator, and organizer of higher-order chromatin structure (see Figure 1A; see Review by R.I. Kumaran et al. on page 929 of this issue). Knockdown of CTCF substantially reduced the association of the cohesin complex with its genomic target loci but not the overall abundance on chromatin, suggesting that CTCF may "focus" chromosome-bound cohesin to specific locations. CTCF and cohesin also colocalize on a key regulatory region in the Kaposi's sarcoma-associated herpesvirus (KSHV) episome during latent infection. Deletion of the CTCF-binding motif from this region abolishes the association of both proteins (Stedman et al., 2008). Conversely, knockdown of the Scc1/Rad21 cohesin subunit had no (Parelho et al., 2008) or a moderate (Wendt et al., 2008) effect on CTCF binding. Although a physical interaction could not be demonstrated, these data show that CTCF is necessary for the correct localization of cohesin to specific sites of the genome during interphase.

Several functional analyses demonstrate that the cohesin complex plays an important role in gene regulation by CTCF. Similar to the removal of CTCF, depletion of Scc1/Rad21 caused activation of KSHV lytic-phase genes in the vicinity of the CTCF-binding site (Stedman et al., 2008). Genome-wide expression profiling in human cells showed significant overlap between the genes that are deregulated upon knockdown of either CTCF or cohesin. These deregulated genes are preferentially located within ~25 kb of a CTCF/cohesion-binding site (Wendt et al., 2008). CTCF has previously been reported to act as an insulator that can block enhancer-promoter interactions. Indeed, cohesin appears to play an essential role in this function (Figure 1B). Knockdown of Scc1/Rad21 or SMC3 resulted in reduced insulator activity of CTCF in reporter assays as well as in the endogenous H19/IGF2 locus (Parelho et al., 2008; Wendt et al., 2008).

Interestingly, CTCF has been reported to mediate pairing between X chromosomes (Xu et al., 2007) and interactions between distant regulatory elements (Splinter et al., 2006). It is tempting to speculate that cohesin, with its unique ability to keep chromatin fibers together, plays a role in the establishment of these higher-order chromatin conformations.

Tying It All Together

The data from yeast, *Drosophila*, and mammalian cells reveal a striking evolutionary plasticity in the targeting mechanisms of cohesin. The distribution of cohesin on chromosomes appears vastly different in these three branches of the evolutionary tree (Figure 1A). It is possible that for sister chromatin cohesion, it may not matter much where exactly cohesin is located, so long as the spacing along the DNA fiber is sufficiently dense. This

very loose constraint may have created opportunities during evolution to re-use the unique molecular properties of cohesin for other regulatory functions. Despite apparently different targeting mechanisms, insulator activity of cohesin has not only been observed in mammals but was also suggested for budding yeast and flies (Dorsett, 2007).

The diversity in location patterns and molecular interactions implies that cohesin has several mechanisms to associate with chromatin. In theory, two fundamentally different modes of DNA association may exist for cohesin: (1) as a tightly closed ring that encircles DNA and (2) via conventional protein-DNA interactions (possibly mediated by other DNA-binding proteins such as CTCF) that do not involve trapping of DNA in a ring. The former mode probably represents a much more stable protein-DNA complex than the latter. In fact, biochemical evidence indicates that both forms may exist (Hirano and Hirano, 2006). Experiments in rat cells using fluorescence recovery after photobleaching (FRAP) also point to the existence of two pools of cohesin during interphase, one pool that is almost irreversibly bound to chromatin, whereas the other displays dynamic exchange that is typical for most chromatin proteins (Gerlich et al., 2006). Interestingly, during G1, nearly all cohesin molecules are in the dynamic state, whereas during G2 both the dynamic and static pools exist. The stably bound pool can also be detected during mitosis, until the onset of anaphase. An attractive interpretation of these FRAP results is that the stable pool represents the cohesin ring embracing the DNA, whereas the dynamic pool may interact with the genome through a different, less stable interaction.

If different modes of interaction also occur in yeast, then the cohesin initially loaded during G1 phase could be dynamically attached via protein-protein or protein-DNA interactions (and thus be easily displaced by the transcription machinery). In G2 it may then become secured as a ring around the DNA, guided by interaction with heterochromatin components. It is also an interesting question whether the CTCF-associated pool of cohesin is dynamically associated or forms a stable attachment as a tightly closed ring around the chromatin fiber. For each of the locations and molecular interactions discussed here, it will be important to know which of these two modes of association occurs. By taking advantage of TEV protease cleavage and other molecular manipulations of the cohesin ring structure, combined with ChIP and FRAP, it should be possible to dissect the different modes of cohesin association with chromatin and its various regulatory roles during interphase.

REFERENCES

Bausch, C., Noone, S., Henry, J.M., Gaudenz, K., Sanderson, B., Seidel, C., and Gerton, J.L. (2007). Mol. Cell. Biol. *27*, 8522–8532.

Bernard, P., Maure, J.F., Partridge, J.F., Genier, S., Javerzat, J.P., and Allshire, R.C. (2001). Science *294*, 2539–2542.

Chang, C.R., Wu, C.S., Hom, Y., and Gartenberg, M.R. (2005). Genes Dev. *19*, 3031–3042.

Dorsett, D. (2007). Chromosoma *116*, 1–13.

Gerlich, D., Koch, B., Dupeux, F., Peters, J.M., and Ellenberg, J. (2006). Curr. Biol. *16*, 1571–1578.

Glynn, E.F., Megee, P.C., Yu, H.G., Mistrot, C., Unal, E., Koshland, D.E., DeRisi, J.L., and Gerton, J.L. (2004). PLoS Biol. *2*, E259. 10.1371/journal.pbio.0020259.

Gullerova, M., and Proudfoot, N.J. (2008). Cell, this issue.

Hirano, M., and Hirano, T. (2006). Mol. Cell *21*, 175–186.

Horsfield, J.A., Anagnostou, S.H., Hu, J.K., Cho, K.H., Geisler, R., Lieschke, G., Crosier, K.E., and Crosier, P.S. (2007). Development *134*, 2639–2649.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Nature *430*, 573–578.

Losada, A. (2007). Chromosoma *116*, 321–329.

Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., Macarthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D., et al. (2008). Chromosoma *117*, 89–102.

Nonaka, N., Kitajima, T., Yokobayashi, S., Xiao, G., Yamamoto, M., Grewal, S.I., and Watanabe, Y. (2002). Nat. Cell Biol. *4*, 89–93.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cell *132*, 422–433.

Pauli, A., Althoff, F., Oliveira, R.A., Heidmann, S., Schuldiner, O., Lehner, C.F., Dickson, B.J., and Nasmyth, K. (2008). Dev. Cell *14*, 239–251.

Schuldiner, O., Berdnik, D., Levy, J.M., Wu, J.S., Luginbuhl, D., Gontang, A.C., and Luo, L. (2008). Dev. Cell *14*, 227–238.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). Genes Dev. *20*, 2349–2354.

Stedman, W., Kang, H., Lin, S., Kissil, J.L., Bartolomei, M.S., and Lieberman, P.M. (2008). EMBO J. *27*, 654–666.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Nature *451*, 796–801.

Xu, N., Donohoe, M.E., Silva, S.S., and Lee, J.T. (2007). Nat. Genet. *39*, 1390–1396.